

# The dawn of artificial intelligence

Powerful computers will reshape humanity's future. How to ensure the promise outweighs the perils



**“T**HE development of full artificial intelligence could spell the end of the human race,” Stephen Hawking warns. Elon Musk fears that the development of artificial intelligence, or AI, may be the biggest existential threat humanity faces. Bill

Gates urges people to beware of it.

Dread that the abominations people create will become their masters, or their executioners, is hardly new. But voiced by a renowned cosmologist, a Silicon Valley entrepreneur and the founder of Microsoft—hardly Luddites—and set against the vast investment in AI by big firms like Google and Microsoft, such fears have taken on new weight. With supercomputers in every pocket and robots looking down on every battlefield, just dismissing them as science fiction seems like self-deception. The question is how to worry wisely.

## You taught me language and...

The first step is to understand what computers can now do and what they are likely to be able to do in the future. Thanks to the rise in processing power and the growing abundance of digitally available data, AI is enjoying a boom in its capabilities (see pages 18-21). Today's “deep learning” systems, by mimicking the layers of neurons in a human brain and crunching vast amounts of data, can teach themselves to perform some tasks, from pattern recognition to translation, almost as well as humans can. As a result, things that once called for a mind—from interpreting pictures to playing the video game “Frogger”—are now within the scope of computer programs. DeepFace, an algorithm unveiled by Facebook in 2014, can recognise individual human faces in images 97% of the time.

Crucially, this capacity is narrow and specific. Today's AI produces the semblance of intelligence through brute number-crunching force, without any great interest in approximating how minds equip humans with autonomy, interests and desires. Computers do not yet have anything approaching the wide, fluid ability to infer, judge and decide that is associated with intelligence in the conventional human sense.

Yet AI is already powerful enough to make a dramatic difference to human life. It can already enhance human endeavour by complementing what people can do. Think of chess, which computers now play better than any person. The best players in the world are not machines however, but what Garry Kasparov, a grandmaster, calls “centaurs”: amalgamated teams of humans and algorithms. Such collectives will become the norm in all sorts of pursuits: supported by AI, doctors will have a vastly augmented ability to spot cancers in medical images; speech-recognition algorithms running on smartphones will bring the internet to many millions of illiterate people in developing countries; digital assistants will suggest promising hypotheses for academic research; image-classification algorithms will allow wearable computers to layer useful information onto people's views of the real world.

Even in the short run, not all the consequences will be posi-

tive. Consider, for instance, the power that AI brings to the apparatus of state security, in both autocracies and democracies. The capacity to monitor billions of conversations and to pick out every citizen from the crowd by his voice or her face poses grave threats to liberty.

And even when there are broad gains for society, many individuals will lose out from AI. The original “computers” were drudges, often women, who performed endless calculations for their higher-ups. Just as transistors took their place, so AI will probably turf out whole regiments of white-collar workers. Education and training will help and the wealth produced with the aid of AI will be spent on new pursuits that generate new jobs. But workers are doomed to dislocations.

Surveillance and dislocations are not, though, what worries Messrs Hawking, Musk and Gates, or what inspires a phalanx of futuristic AI films that Hollywood has recently unleashed onto cinema screens. Their concern is altogether more distant and more apocalyptic: the threat of autonomous machines with superhuman cognitive capacity and interests that conflict with those of *Homo sapiens*.

Such artificially intelligent beings are still a very long way off; indeed, it may never be possible to create them. Despite a century of poking and prodding at the brain, psychologists, neurologists, sociologists and philosophers are still a long way from an understanding of how a mind might be made—or what one is. And the business case for even limited intelligence of the general sort—the sort that has interests and autonomy—is far from clear. A car that drives itself better than its owner sounds like a boon; a car with its own ideas about where to go, less so.

## ...I know how to curse

But even if the prospect of what Mr Hawking calls “full” AI is still distant, it is prudent for societies to plan for how to cope. That is easier than it seems, not least because humans have been creating autonomous entities with superhuman capacities and unaligned interests for some time. Government bureaucracies, markets and armies: all can do things which unaided, unorganised humans cannot. All need autonomy to function, all can take on life of their own and all can do great harm if not set up in a just manner and governed by laws and regulations.

These parallels should comfort the fearful; they also suggest concrete ways for societies to develop AI safely. Just as armies need civilian oversight, markets are regulated and bureaucracies must be transparent and accountable, so AI systems must be open to scrutiny. Because systems designers cannot foresee every set of circumstances, there must also be an off-switch. These constraints can be put in place without compromising progress. From the nuclear bomb to traffic rules, mankind has used technical ingenuity and legal strictures to constrain other powerful innovations.

The spectre of eventually creating an autonomous non-human intelligence is so extraordinary that it risks overshadowing the debate. Yes, there are perils. But they should not obscure the huge benefits from the dawn of AI. ■

VISIT...

LANZAROTE  
*Caliente*.COM



## Rise of the machines

### Artificial intelligence scares people—excessively so

**E**LON MUSK busies himself building other people's futures. A serial entrepreneur who made his first fortune in the early days of the world wide web, he has since helped found a solar-power company to generate green electricity, an electric-car firm to liberate motorists from the internal-combustion engine, and a rocketry business—SpaceX—to pursue his desire to see a human colony on Mars within his lifetime. It makes him the sort of technologist you would expect might look on tomorrow with unbridled optimism.

Not all future technology meets with his approval, though. In a speech in October at the Massachusetts Institute of Technology, Mr Musk described artificial intelligence (AI) as “summoning the demon”, and the creation of a rival to human intelligence as probably the biggest threat facing the world. He is not alone. Nick Bostrom, a philosopher at the University of Oxford who helped develop the notion of “existential risks”—those that threaten humanity in general—counts advanced artificial intelligence as one such, alongside giant asteroid strikes and all-out nuclear war. Lord Rees, who used to run the Royal Society, Britain's foremost scientific body, has since founded the Centre for the Study of Existential Risk, in Cambridge, which takes the risks posed by AI just as seriously.

Such worries are a mirror image of the optimism suffusing the field itself, which has enjoyed rapid progress over the past

couple of years. Firms such as Google, Facebook, Amazon and Baidu have got into an AI arms race, poaching researchers, setting up laboratories and buying start-ups. The insiders are not, by and large, fretting about being surpassed by their creations. Their business is not so much making new sorts of minds as it is removing some of the need for the old sort, by taking tasks that used to be things which only people could do and making them amenable to machines.

The torrent of data thrown off by the world's internet-connected computers, tablets and smartphones, and the huge amounts of computing power now available for processing that torrent, means that their algorithms are more and more capable of understanding languages, recognising images and the like. Business is taking notice. So are those who worry about technology taking away people's jobs. Lots of work depends on recognising patterns and translating symbols. If computers replace some of the people now doing this, either by providing an automated alternative or by making a few such workers far more productive, there will be more white collars in the dole queue.

Signs of the AI boom are everywhere. Last year, Google was rumoured to have paid \$400m for DeepMind, a London-based AI startup. It snatched the firm from under the nose of Facebook, which boasts its own dedicated AI research laboratory,

headed by Yann LeCun, a star researcher hired from New York University. Google once employed Andrew Ng, an AI guru from Stanford University—until Baidu poached him last year to head up a new, Silicon Valley-based lab of its own. Firms such as Narrative Science, in Chicago, which hopes to automate the writing of reports (and which is already used by *Forbes*, a business magazine, to cover basic financial stories), and Kensho, of Cambridge, Massachusetts, which aims to automate some of the work done by “quants” in the financial industry, have been showered in cash by investors. On April 13th IBM announced plans to use a version of its Watson computer—which crushed two puny human champions at an obscurantist American quiz show called *Jeopardy!* in 2011—to analyse health records, looking for medical insights.

### Deep thought

Research into artificial intelligence is as old as computers themselves. Much of the current excitement concerns a subfield of it called “deep learning”, a modern refinement of “machine learning”, in which computers teach themselves tasks by crunching large sets of data. Algorithms created in this manner are a way of bridging a gap that bedevils all AI research: by and large, tasks that are hard for humans are easy for computers, and vice versa. The simplest computer can run rings around the brightest person when it comes to wading through complicated mathematical equations. At the same time, the most powerful computers have, in the past, struggled with things that people find trivial, such as recognising faces, decoding speech and identifying objects in images.

One way of understanding this is that for humans to do things they find difficult, ►►



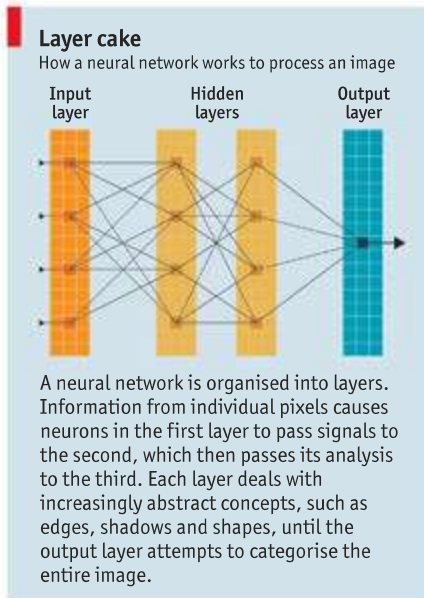
such as solving differential equations, they have to write a set of formal rules. Turning those rules into a program is then pretty simple. For stuff human beings find easy, though, there is no similar need for explicit rules—and trying to create them can be hard. To take one famous example, adults can distinguish pornography from non-pornography. But describing how they do so is almost impossible, as Potter Stewart, an American Supreme Court judge, discovered in 1964. Frustrated by the difficulty of coming up with a legally watertight definition, he threw up his hands and wrote that, although he could not define porn in the abstract, “I know it when I see it.”

Machine learning is a way of getting computers to know things when they see them by producing for themselves the rules their programmers cannot specify. The machines do this with heavy-duty statistical analysis of lots and lots of data.

Many systems use an old and venerable piece of AI technology, the neural network, to develop the statistics that they need. Neural networks were invented in the 1950s by researchers who had the idea that, though they did not know what intelligence was, they did know that brains had it. And brains do their information processing not with transistors, but with neurons. If you could simulate those neurons—spindly, highly interlinked cells that pass electrochemical signals between themselves—then perhaps some sort of intelligent behaviour might emerge.

### Caught by the net

Neurons are immensely complex. Even today, the simulations used in AI are a stick-figure cartoon of the real thing. But early results suggested that even the crudest networks might be good for some tasks. Chris Bishop, an AI researcher with Microsoft, points out that telephone companies have, since the 1960s, been using echo-cancelling algorithms discovered by neural networks. But after such early successes the idea lost its allure. The computing power then available limited the size of the networks that



could be simulated, and this limited the technology's scope.

In the past few years, however, the remarkable number-crunching power of chips developed for the demanding job of drawing video-game graphics has revived interest. Early neural networks were limited to dozens or hundreds of neurons, usually organised as a single layer. The latest, used by the likes of Google, can simulate billions. With that many ersatz neurons available, researchers can afford to take another cue from the brain and organise them in distinct, hierarchical layers (see diagram). It is this use of interlinked layers that puts the “deep” into deep learning.

Each layer of the network deals with a different level of abstraction. To process an image, for example, the lowest layer is fed the raw images. It notes things like the brightness and colours of individual pixels, and how those properties are distributed across the image. The next layer combines these observations into more abstract categories, identifying edges, shadows and the like. The layer after that

will analyse those edges and shadows in turn, looking for combinations that signify features such as eyes, lips and ears. And these can then be combined into a representation of a face—and indeed not just any face, but even a new image of a particular face that the network has seen before.

To make such networks useful, they must first be trained. For the machine to program itself for facial recognition, for instance, it will be presented with a “training set” of thousands of images. Some will contain faces and some will not. Each will be labelled as such by a human. The images act as inputs to the system; the labels (“face” or “not face”) as outputs. The computer's task is to come up with a statistical rule that correlates inputs with the correct outputs. To do that, it will hunt at every level of abstraction for whatever features are common to those images showing faces. Once these correlations are good enough, the machine will be able, reliably, to tell faces from not-faces in its training set. The next step is to let it loose on a fresh set of images, to see if the facial-recognition rules it has extracted hold up in the real world.

By working from the bottom up in this way, machine-learning algorithms learn to recognise features, concepts and categories that humans understand but struggle to define in code. But such algorithms were, for a long time, narrowly specialised. Programs often needed hints from their designers, in the form of hand-crafted bits of code that were specific to the task at hand—one set of tweaks for processing images, say, and another for voice recognition.

Earlier neural networks, moreover, had only a limited appetite for data. Beyond a certain point, feeding them more information did not boost their performance. Modern systems need far less hand-holding and tweaking. They can also make good use of as many data as you are able to throw at them. And because of the internet, there are plenty of data to throw.

Big internet companies like Baidu, Google and Facebook sit on huge quantities of information generated by their users. Reams of e-mails; vast piles of search and buying histories; endless images of faces, cars, cats and almost everything else in the world pile up in their servers. The people who run those firms know that these data contain useful patterns, but the sheer quantity of information is daunting. It is not daunting for machines, though. The problem of information overload turns out to contain its own solution, especially since many of the data come helpfully pre-labelled by the people who created them. Fortified with the right algorithms, computers can use such annotated data to teach themselves to spot useful patterns, rules and categories within.

The results are impressive. In 2014 Facebook unveiled an algorithm called DeepFace that can recognise specific human

### Some you win, some you lose

Image-recognition software's analysis of what a picture represents



“A person riding a motorcycle on a dirt road”



“A yellow school bus parked in a car park”

Source: “Show and Tell: A Neural Image Caption Generator”, Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan

**Spot...**

This is a Cardigan Corgi, one of two types of Welsh Corgi. Slightly larger and heavier-boned than the Pembroke Corgi.

► faces in images around 97% of the time, even when those faces are partly hidden or poorly lit. That is on a par with what people can do. Microsoft likes to boast that the object-recognition software it is developing for Cortana, a digital personal assistant, can tell its users the difference between a picture of a Pembroke Welsh Corgi and a Cardigan Welsh Corgi, two dog breeds that look almost identical (see pictures above). Some countries, including Britain, already use face-recognition technology for border control. And a system capable of recognising individuals from video footage has obvious appeal for policemen and spies. A report published on May 5th showed how America's spies use voice-recognition software to convert phone calls into text, in order to make their contents easier to search.

But, although the internet is a vast data trove, it is not a bottomless one. The sorts of human-labelled data that machine-learning algorithms thrive on are a finite resource. For this reason, a race is on to develop "unsupervised-learning" algorithms, which can learn without the need for human help.

There has already been lots of progress. In 2012 a team at Google led by Dr Ng showed an unsupervised-learning machine millions of YouTube video images. The machine learned to categorise common things it saw, including human faces and (to the amusement of the internet's denizens) the cats—sleeping, jumping or skateboarding—that are ubiquitous online. No human being had tagged the videos as containing "faces" or "cats". Instead, after seeing zillions of examples of each, the machine had simply decided that the statistical patterns they represented were common enough to make into a category of object.

The next step up from recognising individual objects is to recognise lots of different ones. A paper published by Andrej Karpathy and Li Fei-Fei at Stanford University describes a computer-vision system that is able to label specific parts of a given picture. Show it a breakfast table, for instance, and it will identify the fork, the banana slices, the cup of coffee, the flowers on the

table and the table itself. It will even generate descriptions, in natural English, of the scene (see picture below)—though the technology is not yet perfect (see picture on previous page).

Big internet firms such as Google are interested in this kind of work because it can directly affect their bottom lines. Better image classifiers should improve the ability of search engines to find what their users are looking for. In the longer run, the technology could find other, more transformative uses. Being able to break down and interpret a scene would be useful for robotics researchers, for instance, helping their creations—from industrial helpmeets to self-driving cars to battlefield robots—to navigate the cluttered real world.

Image classification is also an enabling technology for "augmented reality", in which wearable computers, such as Google's Glass or Microsoft's HoloLens, overlay useful information on top of the real world. Enlitic, a firm based in San Francisco, hopes to employ image recognition to analyse x-rays and MRI scans, looking for problems that human doctors might miss.

And deep learning is not restricted to images. It is a general-purpose pattern-recognition technique, which means, in principle, that any activity which has access to large amounts of data—from running an insurance business to research into genetics—might find it useful. At a recent competition held at CERN, the world's biggest particle-physics laboratory, deep-learning algorithms did a better job of spotting the signatures of subatomic particles than the software written by physicists—even though the programmers who created these algorithms had no particular knowledge of physics. More whimsically, a group of researchers have written a program that

**...the difference**

This is a Pembroke Corgi, the more common of the two types of Welsh Corgi and the one favoured by the queen.

learnt to play video games such as "Space Invaders" better than people can.

Machine translation, too, will be improved by deep learning. It already uses neural networks, benefiting from the large quantity of text available online in multiple languages. Dr Ng, now at Baidu, thinks good speech-recognition programs running on smartphones could bring the internet to many people in China who are illiterate, and thus struggle with ordinary computers. At the moment, 10% of the firm's searches are conducted by voice. He believes that could rise to 50% by 2020.

And those different sorts of AI can be linked together to form an even more capable system. In May 2014, for instance, at a conference in California, Microsoft demonstrated a computer program capable of real-time translation of spoken language. The firm had one of its researchers speak, in English, to a colleague in Germany. This colleague heard her interlocutor speaking in German. One AI program decoded sound waves into English phrases. Another translated those phrases from English into German, and a third rendered them into German speech. The firm hopes, one day, to build the technology into Skype, its internet-telephony service.

**No ghost in the machine**

Better smartphones, fancier robots and bringing the internet to the illiterate would all be good things. But do they justify the existential worries of Mr Musk and others? Might pattern-recognising, self-programming computers be an early, but crucial, step on the road to machines that are more intelligent than their creators?

The doom-mongers have one important fact on their side. There is no result from decades of neuroscientific research to suggest that the brain is anything other than a machine, made of ordinary atoms, employing ordinary forces and obeying the ordinary laws of nature. There is no mysterious "vital spark", in other words, that is necessary to make it go. This suggests that building an artificial brain—or even a machine that looks different from a ►

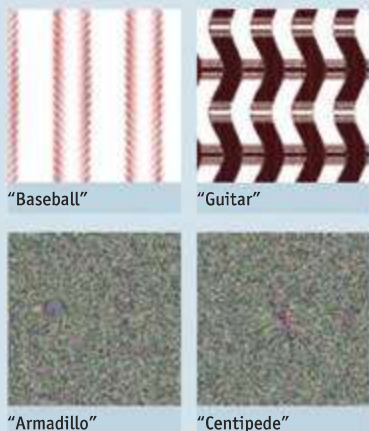


A computer's take on breakfast. Each box is part of a picture that it has identified and labelled, including "bouquet of red flowers", "glass of water with ice and lemon" and "banana slices". The whole was deemed to show a "dining table with breakfast items".



### Pattern recognition

A few examples of the machine getting it spectacularly wrong



Source: "Show and Tell: A Neural Image Caption Generator", Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan

► brain but does the same sort of thing—is possible in principle.

But doing something in principle and doing it in fact are not remotely the same thing. Part of the problem, says Rodney Brooks, who was one of AI's pioneers and who now works at Rethink Robotics, a firm in Boston, is a confusion around the word "intelligence". Computers can now do some narrowly defined tasks which only human brains could manage in the past (the original "computers", after all, were humans, usually women, employed to do the sort of tricky arithmetic that the digital sort find trivially easy). An image classifier may be spookily accurate, but it has no goals, no motivations, and is no more conscious of its own existence than is a spreadsheet or a climate model. Nor, if you were trying to recreate a brain's workings, would you necessarily start by doing the things AI does at the moment in the way that it now does them. AI uses a lot of brute force to get intelligent-seeming responses from systems that, though bigger and more powerful now than before, are no more like minds than they ever were. It does not seek to build systems that resemble biological minds. As Edsger Dijkstra, another pioneer of AI, once remarked, asking whether a computer can think is a bit like asking "whether submarines can swim".

### A snare and an illusion

Nothing makes this clearer than the ways in which AI programs can be spoofed. A paper to be presented at a computer-vision conference in June shows optical illusions designed to fool image-recognition algorithms (see picture above). These offer insight into how the algorithms operate—by matching patterns to other patterns, but doing so blindly, with no recourse to the sort of context (like realising a baseball is a physical object, not just an abstract pattern vaguely reminiscent of stitching) that stops

people falling into the same traps. It is even possible to construct images that, to a human, look like meaningless television static, but which neural networks nevertheless confidently classify as real objects.

This is not to say that progress in AI will have no unpleasant consequences, at least for some people. And, unlike previous waves of technological change, quite a few of those people may be middle class. Take Microsoft's real-time translation. The technology it demonstrated was far from perfect. No one would mistake its computer-translated speech for the professionally translated sort. But it is adequate to convey the gist of what is being said. It is also cheaper and more convenient than hiring a human interpreter. Such an algorithm could therefore make a limited version of what is presently a costly, bespoke service available to anyone with a Skype account. That might be bad for interpreters. But it would be a boon for everyone else. And Microsoft's program will only get better.

The worry that AI could do to white-collar jobs what steam power did to blue-collar ones during the Industrial Revolution is therefore worth taking seriously. Examples, such as Narrative Science's digital financial journalist and Kensho's quant, abound. Kensho's system is designed to interpret natural-language search queries such as, "What happens to car firms' share prices if oil drops by \$5 a barrel?" It will then scour financial reports, company filings, historical market data and the like, and return replies, also in natural language, in seconds. The firm plans to offer the software to big banks and sophisticated traders. Yseop, a French firm, uses its natural-language software to interpret queries, chug through data looking for answers, and then write them up in English, Spanish, French or German at 3,000 pages a second. Firms such as L'Oréal and VetOnline.com already use it for customer support on their websites.

Nor is this just a theoretical worry, for some white-collar jobs are already being

lost to machines. Many firms use computers to answer telephones, for instance. For all their maddening limitations, and the need for human backup when they encounter a query they cannot understand, they are cheaper than human beings. Forecasting how many more jobs might go the same way is much harder—although a paper from the Oxford Martin School, published in 2013, scared plenty of people by concluding that up to half of the job categories tracked by American statisticians might be vulnerable.

Technology, though, gives as well as taking away. Automated, cheap translation is surely useful. Having an untiring, lightning-fast computer checking medical images would be as well. Perhaps the best way to think about AI is to see it as simply the latest in a long line of cognitive enhancements that humans have invented to augment the abilities of their brains. It is a high-tech relative of technologies like paper, which provides a portable, reliable memory, or the abacus, which aids mental arithmetic. Just as the printing press put scribes out of business, high-quality AI will cost jobs. But it will enhance the abilities of those whose jobs it does not replace, giving everyone access to mental skills possessed at present by only a few. These days, anyone with a smartphone has the equivalent of a city-full of old-style human "computers" in his pocket, all of them working for nothing more than the cost of charging the battery. In the future, they might have translators or diagnosticians at their beck and call as well.

Cleverer computers, then, could be a truly transformative technology, though not—at least, not yet—for the reasons given by Mr Musk or Lord Rees. One day, perhaps, something like the sort of broad intelligence that characterises the human brain may be recreated in a machine. But for now, the best advice is to ignore the threat of computers taking over the world—and check that they are not going to take over your job first. ■

